Chapter 3

Beam Combination and Fringe Measurement

Wesley A. Traub

Smithsonian Astrophysical Observatory
Cambridge, Massachusetts

This chapter discusses how to calculate the interference of wavefronts for telescopes and interferometers, how to separate astrophysical effects from instrumental effects, and how to optimize visibility measurements.

We focus on homodyne detection, a technique in which stellar wavefront segments are combined with each other in real time to make an image or a fringe pattern, whose intensity is then detected.

The alternative is heterodyne detection, a technique in which each stellar wavefront segment is combined with a local oscillator signal to make a beat frequency which is recorded, and at a later time a collection of simultaneously recorded beat signals is combined to form an image or a fringe pattern. The heterodyne technique and its applicability to wavelengths of about 10 μm and longer is discussed by Townes in Chapter 4.

However, despite their different detection techniques, homodyne and heterodyne interferometry still share identical basic principles, and much of the material in this chapter applies equally to both.

3.1 Elements of Beam Combination

In this section we present simple examples of beam combination, for one- and two-aperture configurations and for three basic types of sources. We distinguish multiple apertures (this section) from multiple telescopes (the following section). From the discussion it should be clear how to generalize each result to more realistic or complex situations. In simple cases it is relatively straightforward to see how to invert observed data to infer source properties, but in general the van Cittert–Zernike theorem is needed, as discussed at the end of this section.

3.1.1 Single Aperture and Point Source

It is instructive to start thinking about interferometers by first considering the case of a single telescope. In fact, an ideal single telescope is also an ideal interferometer. If it were possible to build a single telescope with a diameter as large as the baseline of an interferometer, we would never even consider building an interferometer, since interferometers are intrinsically more complex than telescopes. Also, many of the technical aspects of interferometers derive directly from the need to duplicate the action of a single large mirror, so it is useful to understand how a single collecting element works.

The simplest possible case is that of a plane wave incident on a single aperture, as shown in Figure 3.1. The collimating element is drawn as a lens, but any equivalent optical system, such as a Cassegrain telescope, will do. The aperture is stopped to a diameter D by a baffle in the (x, y) plane. The symmetry axis (z) in Figure 3.1 is a line from the center of the star through the optical center of the telescope lens.

Throughout this chapter we will assume that the optical systems are completely free of geometrical optical aberrations. For a useful book on telescope optics see Schroeder (2000).

The incident light from a real star is a stream of photons arriving at random times from a range of random angles within the angular diameter of the star. A single photon from this stream effectively exists over an arbitrarily large area on the surface of a sphere centered on its emitting atom, prior to detection. The photon simultaneously senses the presence of all the details of the collecting aperture, which can be of arbitrary shape, size, and degree of topological connectedness.

It helps at this point to forget about photons and think instead about waves. It also helps to think about a wave as a series of little spherical waves which repeatedly generate themselves at all points across the wavefront, and propagate outwards, but which in free space only end up propagating in the forward direction because this is the only direction in which the little waves constructively interfere. These are called Huygens' wavelets in optics texts, where the term "wavelet" means little wave. It does not mean a wave packet in the modern mathematical sense of the word "wavelet"!

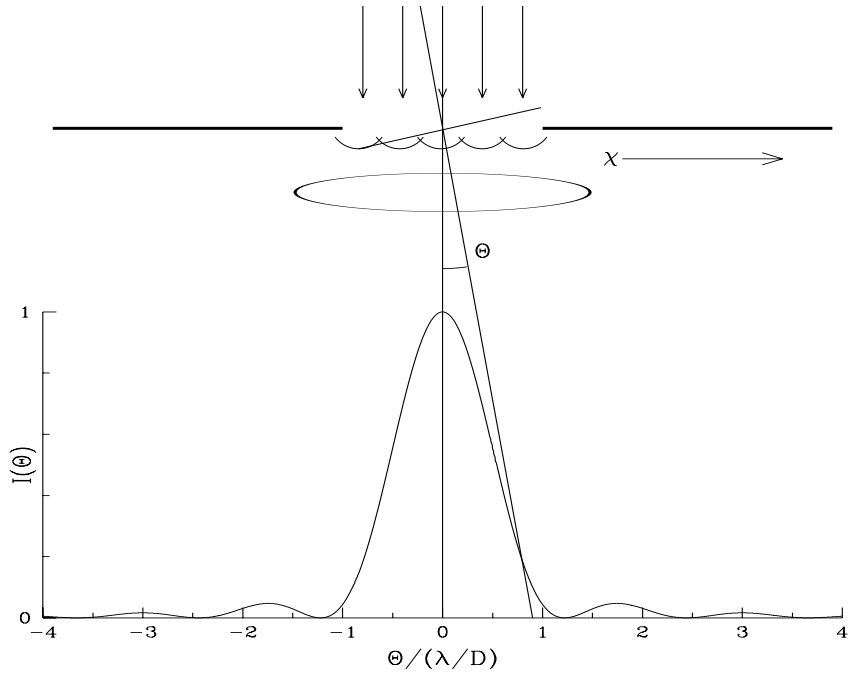

Figure 3.1: Delta-function star and single telescope. Input wavefronts arrive at the aperture from angle $\theta = 0$ only, but output wavefronts exist at all angles, albeit concentrated near the input angle. The intensity pattern plotted here is the circular aperture diffraction pattern $I_{\rm tel}(\theta) = [2J_1(\pi\theta D/\lambda)/(\pi\theta D/\lambda)]^2$.

The incident idealized photon is also monochromatic, and therefore has essentially infinite extent in the direction of propagation. The corresponding classical wave has the same extent. If we think of an emitting atom at the surface of the star where the photon originated as a classical oscillator, then we may define a wavefront from this oscillator as the collection of all points on the outwardly propagating wave which were generated at the same time. There are therefore an infinite series of concentric wavefronts being emitted over time, and we may choose any one of them to consider as the wave moves through our optical system. I find it useful to think of the wavefront as the surface on which the wave has its maximum positive electric field strength. Since the wave is periodic, successive wavefronts are separated by one wavelength.

As the wavefront crosses the plane of the entrance pupil, the wavelets in the center of the pupil continue as before, propagating a plane wavefront. However at the edge of the pupil there is no longer a reinforcement of wavelets from the part of the incident wavefront that is now blocked, and the wavelets inside the pupil will start to spread transversely into the geometric optics shadow region.

For an alternative mental model of wavefront propagation, one could also think of row after row of people marching shoulder to shoulder in a huge parade, with the people in

the middle being kept to the forward direction by their neighbor's shoulders. When the marching rows encounter a wide gate open in a long wall, the people in the center are still constrained to march forward by their neighbors, but the few folks nearest the gate edges will feel unconstrained on one side, and without guidance their paths will begin to divert from that of the main group.

Actually the row-of-people analogy is not quite as far fetched as it sounds, because in principle each of the marching people has a de Broglie wavelength given by $\lambda = h/mc$, so to the extent that these people are mindless particles with no other influences on them, they will behave like diffracting particles. More practically, a beam of electrons, neutrons, protons, atoms, or molecules will follow the same rules, as has been demonstrated many times.

In astronomical telescopes we are only interested in what happens to the wavelets at large distances (compared to a wavelength) from the diffracting edges, i.e., in the far-field diffraction pattern. Fortunately for us, this is relatively simple to calculate, given that each wavelet has a known amplitude and phase, and given that we have a rule for combining electric field amplitudes and another rule for converting the total amplitude to an intensity, which is the measurable quantity of interest.

Another measurable quantity is the state of polarization of the electric field of the detected photon. In general we will ignore polarization because the net effect of polarization on diffraction is most often a second-order effect. There is one important exception, however, in the case of reflection from mirrors, which we will discuss later in this chapter.

In Figure 3.1, we show an aperture followed by a lens and a focal plane. According to geometric optics, the star will be focussed in the focal plane, and the image will be a perfect replica of the star, here a delta-function. However according to physical optics we must find the image by adding up the amplitudes of the wavelets which make it through the aperture and propagate in the direction of any point in the focal plane.

Recall that an ideal lens merely acts to convert an input direction of propagation into an output location in the focal plane. A pinhole does the same thing, but a lens is better in the sense that it converts an incident ray anywhere on its surface into a ray heading for a single position in the focal plane. It is also helpful to recall that the ray through the center of the lens is not deviated, so input direction equals output direction. Thus in Figure 3.1 we show intensity in the focal plane in terms of an output angle θ. This is equivalent to position in the focal plane, divided by the focal length of the imaging optics.

At a given point θ in the focal plane, the total amplitude is the sum of all the wavelet amplitudes at the aperture heading in the direction θ, allowing for their relative phases. From Maxwell's equations we know that the electric field amplitude is sinusoidal in space and time, so that in one dimension and for one polarization, the electric field amplitude E can be written conveniently as the real part of the complex amplitude $A(z,t) = e^{i(\omega t - kz + \phi)}$. We use conventional notation where z is linear distance in the direction of propagation, t

is time, $k = 2\pi/\lambda = nk_0 = 2\pi n/\lambda_0$ is the magnitude of the wave vector (or propagation vector), n is the index of refraction, nz is the "optical path," $\lambda = \lambda_0/n = c/n\nu$ is the wavelength in the medium, ν is the temporal frequency of oscillation, $\omega = 2\pi\nu$ is the angular frequency, and ϕ is the phase of the wave (Born and Wolf, 1999).

Let us define the position of measurement of the amplitude to be $z = 0$, the time of measurement to be $t = 0$, and the medium of propagation to be vacuum, so $n = 1$. This reduces the complex amplitude to $e^{i\phi}$.

At first sight it may appear that we have defined away all the interesting physics, but the seemingly ridiculous simplicity of $e^{i\phi}$ is in reality the heart of the problem of calculating interference effects. We have merely stripped away the non-essential parts.

From a quantum-mechanical point of view, the $e^{i\phi}$ term is a propagator of a probability amplitude from one place and time to another, where ϕ represents the change in phase (modulo 2π) of the probability amplitude along the minimum path.

The phase is calculated across a tilted surface in the pupil, oriented at an angle θ with respect to the incoming wavefront. There are an infinite number of such tilted surfaces. The relative strength of an outgoing wavefront parallel to one of these surfaces is determined by adding up all the wavelets on that surface. The phase at each point is 2π times the distance between the input and output wavefronts, in units of wavelength. Let us focus on the aperture's x-dimension for the moment. The phase of a wavelet is

$$\begin{aligned} \phi(z) &= 2\pi x \sin(\theta)/\lambda & (3.1)\\ &\simeq 2\pi x\theta/\lambda & (3.2) \end{aligned}$$

where $x\sin(\theta)$is the distance between the incoming wavefront from direction $\theta = 0$ and the outgoing wavefront at angle θ, and we assume $\theta \ll 1$.

The net output amplitude from the telescope in the direction θ is $A_{\rm tel}(\theta)$, which we calculate as the algebraic sum of all wavelets across the pupil.

$$\begin{aligned} A_{\rm tel}(\theta) &= \sum(\text{wavelets}) & (3.3)\\ &= \int_{\rm pupil} e^{i\phi(x)}dx & (3.4)\\ &= \int_{-D/2}^{+D/2} e^{i(2\pi x\theta/\lambda)}dx & (3.5)\\ &= \frac{\lambda}{2\pi i\theta}\left[e^{+i\pi\theta D/\lambda} - e^{-i\pi\theta D/\lambda}\right] & (3.6)\\ &= \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda}D & (3.7) \end{aligned}$$

The measured intensity I is the squared magnitude of the amplitude.

$$I_{\rm tel}(\theta) = |A_{\rm tel}|^2 \quad (3.8)$$

$$I_{\rm tel}(\theta) = \left[\frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda}\right]^2 D^2 \tag{3.9}$$

The intensity pattern is thus a $\mathrm{sinc}(X) \equiv \sin(X)/X$ function, with a strong central peak and small secondary peaks, as shown in Figure 3.1. The first zero is the solution of $I_{\rm tel}(\theta_{\rm tel}) = 0$ and is given by

$$\theta_{\rm tel} = \lambda/D. \tag{3.10}$$

The corresponding result for a two-dimensional circular aperture is found by replacing $\int_D$ by $\int_{\rm circle}$ with the result

$$I_{\rm tel}(\theta) = \left[\frac{2J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda}\right]^2 D^2 \tag{3.11}$$

where $J_1(X)$ is the Bessel function of first order, roughly similar to a damped sine function. Numerically $J_1(X)$ for any real X can be calculated using the BESSJ1 routine in Numerical Recipes (Press *et al.*, 1992). The first zero-intensity angle is the solution of $I_{\rm tel}(\theta_{\rm tel}) = 0$ and is given by $X = 1.22\pi$ or

$$\theta_{\rm tel} = 1.22\lambda/D \tag{3.12}$$

which is the famous relation for an unobstructed circular aperture. The full-width at half-maximum (FWHM) of the intensity pattern is roughly approximated by the value of $\theta_{\rm tel}$, so this value is often loosely referred to as the diameter of the diffraction-limited image.

Here are several useful variations on the same theme.

Constant Phase

Suppose we add a constant phase ϕ_0 across the aperture. Then we get

$$A_{\rm tel}(\theta) = \int_{-D/2}^{+D/2} e^{i(2\pi x\theta/\lambda+\phi_0)}dx \tag{3.13}$$

$$= \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} e^{i\phi_0} D \tag{3.14}$$

and $I_{\rm tel}$ is unchanged.

Star Off-Axis

Suppose that the star moves off the telescope axis, or equivalently that the telescope is pointed away from the star by an angle θ_0. The input wavefronts are then tilted by θ_0, and the summing of phases, determined by the distance from the input and output wavefronts, yields

$$A_{\rm tel}(\theta) = \int_{-D/2}^{+D/2} e^{i(2\pi x(\theta-\theta_0)/\lambda}dx \tag{3.15}$$

$$= \frac{\sin(\pi(\theta-\theta_0)D/\lambda)}{\pi(\theta-\theta_0)D/\lambda} D \tag{3.16}$$

and we see that the intensity pattern will be shifted to a new center at position θ_0 in the focal plane, just as one would expect from geometric optics.

Phase Step: Speckle

Suppose we add a phase step of π across half of the aperture. This simulates the action of the turbulent atmosphere in a very simple case. The net amplitude with this phase step is

$$A_{\rm tel}(\theta) = \int_0^{+D/2} e^{i(2\pi x\theta/\lambda+\pi/2)}dx + \int_{-D/2}^{0} e^{i(2\pi x\theta/\lambda-\pi/2)}dx \quad (3.17)$$

$$= -\frac{\sin^2(\pi\theta D/2\lambda)}{\pi\theta D/2\lambda}D \quad (3.18)$$

and the intensity is

$$I_{\rm tel}(\theta) = \left[\frac{\sin^2(\pi\theta D/2\lambda)}{\pi\theta D/2\lambda}\right]^2 D^2. \quad (3.19)$$

This pattern has two main peaks offset from the axis by about $\pm\lambda/D$, and with widths about λ/D, plus small secondary peaks. The effect of the phase step is to split the unperturbed image into two pieces, each of which looks rather similar to the original image. The perturbed images are called *speckles*. If more phase steps are added, more speckles will appear, but each will still be a more or less faithful copy of the diffraction-limited unperturbed case. This is the basis for speckle interferometry. One of the simplest techniques to recover the original image from a speckle pattern is called "shift and add," which seeks to superpose the speckles on a common axis, post detection. From the example here, one can see why this technique enjoys some success.

Central Obscuration

Suppose that the telescope aperture has an outer width D and a central obscuration of inner width d. In the one-dimensional case, the electric field amplitude in the focal plane is then

$$A_{\rm tel}(\theta) = \int_{-D/2}^{-d/2} e^{i(2\pi x\theta/\lambda)}dx + \int_{+d/2}^{+D/2} e^{i(2\pi x\theta/\lambda)}dx \quad (3.20)$$

$$= \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda}D - \frac{\sin(\pi\theta d/\lambda)}{\pi\theta d/\lambda}d \quad (3.21)$$

and the light intensity is given by

$$I_{\rm tel}(\theta) = \left[\frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda}D - \frac{\sin(\pi\theta d/\lambda)}{\pi\theta d/\lambda}d\right]^2. \quad (3.22)$$

By analogy, the corresponding expression for the intensity from a two-dimensional circular aperture of diameter D, with central obscuration (e.g., a secondary mirror) of diameter d is given by

$$I_{\rm tel}(\theta) = \left[\frac{2J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda}D - \frac{2J_1(\pi\theta d/\lambda)}{\pi\theta d/\lambda}d\right]^2. \tag{3.23}$$

Comparing this with the non-obscured case, we see that for a telescope primary mirror of a given diameter D and a finite secondary diameter d, the intensity distribution has a slightly narrower central core (i.e., slightly better angular resolution) but at the expense of significantly stronger diffraction rings around the central core. The reason that the core is narrower is that the inner portion of the aperture, i.e., the low angular resolution part, has been removed, leaving the outer portions which are responsible for the high angular resolution performance. The sidelobes increase for the reason that, at angles outside the central core, the missing central portion means that there are fewer central wavelets available to provide phase cancellation with the edge wavelets. Thus more of the edge wavelet power appears in an aliased form outside the central core.

Some additional interesting possibilities that could be calculated include tapering the transmission factor of the pupil near the edges so as to *apodise* ("remove the feet"), or reduce, the secondary diffraction rings, or adding a phase screen across the aperture in order to delay the phase near the outer edges and thereby alter the phase of the diffraction rings.

With these examples worked out, we can now graduate quickly to several other key cases of interest to interferometry.

3.1.2 Two Apertures and Point Source

The two-aperture calculation proceeds as with the one-aperture case, but with the single opening replaced by two openings of equal diameter D, separated by a baseline B, as shown in Figure 3.2. The configuration is that of a pointed interferometer, and can be achieved by masking the primary mirror of a large telescope as indicated in the Figure; this is nominally what Michelson did at the 100-inch telescope, and it also describes the original MMT mirrors, and the current LBT. In a very large interferometer, which is too large to mount on a single pointed platform, one can literally join the focal planes of two or more separate telescopes, in which case additional issues of path length, magnification, field rotation, and polarization must be considered.

For two apertures the amplitude $A_{\rm int}$ of the electric field in the focal plane of a two-element interferometer is given by

$$\begin{aligned} A_{\rm int}(\theta) &= \sum(\text{wavelets}) && (3.24)\\ &= \int_{\rm pupil} e^{i\phi(x)}dx && (3.25)\\ &= \int_{+B/2-D/2}^{+B/2+D/2} e^{i(2\pi x\theta/\lambda)}dx + \int_{-B/2-D/2}^{-B/2+D/2} e^{i(2\pi x\theta/\lambda)}dx && (3.26)\\ &= \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda}\cos(\pi\theta B/\lambda)2D && (3.27) \end{aligned}$$

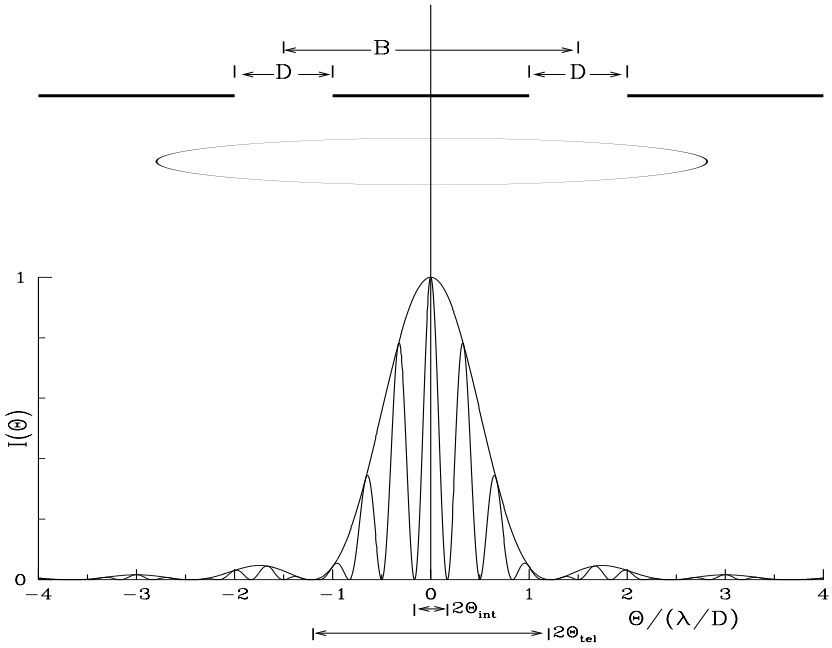

Figure 3.2: Two apertures and a single delta-function source. The two apertures form a pair of superposed diffraction-limited images, crossed by interference fringes. The figure is drawn for the case $B/D = 3$, so the expected number of fringes in the packet is $N_{\rm packet} = 7.3$, which is about what is seen here.

and the intensity is given by

$$I_{\rm int}(\theta) = 2I_{\rm tel}(\theta)\,[1 + \cos(2\pi\theta B/\lambda)] \tag{3.28}$$

which is the product of two terms, the broad envelope of a single-telescope diffraction pattern, and the rapidly varying interference term which depends only on the distance between the telescopes, as shown in Figure 3.2. Clearly the $I_{\rm tel}$ term can be either the one-dimensional one shown here, or a two-dimensional J_1 function derived above.

The first zero of the intensity pattern is the solution of $I_{\rm int}(\theta_{\rm int}) = 0$ and is given by

$$\theta_{\rm int} = \lambda/2B \tag{3.29}$$

which is also the width (FWHM) of one of the narrow fringes, and therefore the angular resolution limit of the interferometer.

Let us define a *fringe packet* as the central lobe of the telescope diffraction pattern. The envelope of the fringe packet has an angular width of $2\theta_{\rm tel}$ between first zeros. The number of fringes $N_{\rm packet}$ in a fringe packet is given by

$$N_{\rm packet} = 2.44B/D. \tag{3.30}$$

3.1.3 Two Apertures and Binary Star

Suppose we have a binary system comprising two equal-magnitude stars separated by $\theta_{\rm bin}$, centered on the axis of a two-aperture interferometer. The amplitudes and intensities from each star must be treated independently, because the sources are not coherent (photons from one star have no knowledge of photons from the other star). In the focal plane we then have two independent interferometer intensity patterns which add to give a binary-star intensity

$$I_{\rm bin} = I_{\rm int}(\theta - \theta_{\rm bin}/2) + I_{\rm int}(\theta + \theta_{\rm bin}/2) \tag{3.31}$$

If the separation is small compared to the width of the fringe packet, i.e., $\theta_{\rm bin} \ll 2\theta_{\rm tel}$, then we can factor out the envelope shape and find

$$\begin{aligned} I_{\rm bin} &\simeq 2I_{\rm tel}(\theta)\left[1 + \cos(2\pi(\theta + \theta_{\rm bin}/2)B/\lambda) + 1 + \cos(2\pi(\theta - \theta_{\rm bin}/2)B/\lambda)\right] &(3.32)\\ &= 4I_{\rm tel}(\theta)\left[1 + V_{\rm bin}\cos(2\pi\theta B/\lambda)\right]. &(3.33)\end{aligned}$$

The coefficient of the interference modulation term is known as the *fringe visibility*, or simply the *visibility*, and is given in this case by

$$V_{\rm bin} = \cos(\pi\theta_{\rm bin}B/\lambda). \tag{3.34}$$

The visibility has its first zero $V_{\rm bin} = 0$ when the binary separation is

$$\theta_{\rm bin} = \lambda/2B. \tag{3.35}$$

The generalization of these results to unequal magnitudes and a two-dimensional configuration is straightforward but messy.

3.1.4 Two Apertures and Uniform Disk

A real star has a finite diameter, and each of the photons emitted from its surface is independent of all other photons (unless there is maser activity taking place, as does happen in the atmospheres of some stars, under appropriate conditions). As with the binary star case, the intensity in the focal plane of a telescope or interferometer is then given by the superposition of appropriately shifted and scaled intensity patterns. For a uniformly bright disk in one-dimension (UD1) whose width is $\theta_{\rm UD1}$, and where we assume that the disk is small compared to the fringe packet width ($\theta_{\rm UD1} \ll 2\theta_{\rm tel}$), we add up the incoherent fringe patterns as follows.

$$\begin{aligned} I_{\rm UD1}(\theta) &= \sum_{\rm disk}(\text{intensities}) &(3.36)\\ &= \int_{\rm disk} I_{\rm int}(\theta - \theta_x)d\theta_x &(3.37)\\ &= \int_{-\theta_{\rm UD1}/2}^{+\theta_{\rm UD1}/2} 2I_{\rm tel}(\theta - \theta_x)\left[1 + \cos(2\pi(\theta - \theta_x)B/\lambda)\right]d\theta_x &(3.38)\\ &\simeq 2I_{\rm tel}(\theta)\left[1 + V_{\rm UD1}\cos(2\pi\theta B/\lambda)\right] &(3.39)\end{aligned}$$

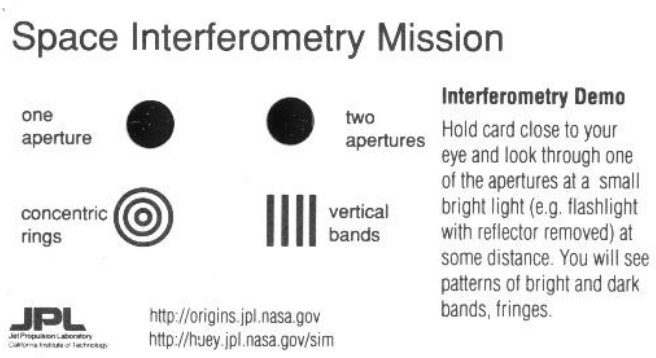

Figure 3.3: The Space Interferometry Mission (SIM) pocket interferometer card. The apertures on the actual card are a single circular hole (.) in a piece of black film on the "one aperture" side, and two holes (..) on the "two apertures" side. With one aperture, a point source will appear as concentric rings, as shown, but with the outer rings progressively fainter than the center one. With two apertures, the central bright peak is crossed by strong vertical modulation bands, as is the secondary ring, and (faintly) the third ring.

Here the fringe visibility of a uniform disk in one-dimension is given by

$$V_{\rm UD1} = \frac{\sin(\pi B\theta_{\rm UD1}/\lambda)}{\pi B\theta_{\rm UD1}/\lambda}. \tag{3.40}$$

By analogy we immediately see that the intensity pattern for a two-dimensional round uniform disk (UD) is given by a similar equation where the visibility is

$$V_{\rm UD} = \frac{2J_1(\pi B\theta_{\rm UD}/\lambda)}{\pi B\theta_{\rm UD}/\lambda}. \tag{3.41}$$

The visibility has its first zero $V_{\rm UD} = 0$ when the star diameter is

$$\theta_{\rm UD} = 1.22\lambda/2B. \tag{3.42}$$

Note that $2J_1(X)$ is similar to $\sin(X)$ in that the central lobe is positive, the first secondary lobe is negative, the second secondary lobe is positive, and so on. In the fringe pattern, this means that the fringes in the alternate lobes have their signs inverted with respect to the extrapolated fringes from neighboring lobes. In other words, counting the central lobe as number 0, the even lobes have phase = 0, and the odd lobes have phase = π.

3.1.5 The Pocket Interferometer

The SIM pocket demonstration card shown in Figure 3.3 is an excellent one-dimensional, two-aperture, mask, which when held close to your eye provides a complete interferometer. Here the telescope is your eye lens, and the focal plane is your retina. The diameter of each aperture is $D \simeq 0.07$ mm, so the telescope diffraction pattern has a width $\theta_{\rm tel} = 1.22\lambda/D \simeq$

2000 arcsec, which is about the angular diameter of the sun or moon. The separation of the apertures is $B \simeq 0.25$ mm, so the angular width of a fringe with this interferometer is $\theta_{\rm int} = \lambda/2B \simeq 200$ arcsec, which is about the width of a Mag-Lite™ filament at a distance of roughly a foot (note however that the filament is much longer than it is wide). The number of fringes in a fringe packet (the angular width of the single-telescope diffraction pattern) is $N_{\rm packet} = 2.44B/D \simeq 8$ in this case, independent of the light source, of course.

The Mag-Lite should be used with the normally-present flashlight reflector removed, so that you see just the filament itself. The Mag-Lite will appear to be a small, unresolved star if it is viewed at arm's length or farther, with the filament rotated (i.e., the flashlight rolled around your line of sight) so that its narrow dimension is parallel to the baseline $\vec{B}$ of the pocket interferometer, thus making a one-dimensional system. In this case you will see a central bright lobe crossed by about eight fringes, oriented perpendicular to the baseline. You will also see the same straight fringes crossing the first and second side lobes of the single-telescope diffraction pattern.

The Mag-Lite can be made into a large-diameter star by rolling it 90° about the line of sight, so that the long axis of the filament is now parallel to $\vec{B}$. In this case you will see a central lobe which is smooth, with no fringes. What has happened is that multiple fringe patterns are now superposed with a range of shifts, and the fringe pattern is washed out, as predicted by the equations above.

3.1.6 Two-Aperture Beam Pattern on Sky

If you were to think like a radio astronomer, you would imagine the antenna pattern to be projected out from the receiver horn of each antenna and thence from the array as a whole and onto the sky. As you move the antenna, or change the phase at an array element, the pattern sweeps across the sky. The received signal is the convolution of the moving pattern and the sources in the sky. A sinusoidal pattern projects out the Fourier component of that spacing of fringes on the object, and therefore the Fourier component of the intensity distribution across the sky.

You can see from this view that if all possible fringe spacings and orientations could be swept across the object, and if their phases (i.e., the relative locations of the central fringe peaks) could be recorded, then these measured quantities would essentially fill the two-dimensional Fourier plane with complex values (an amplitude and phase pair). A Fourier transformation of these values would then yield a perfect image of the source. This is not only a good mental picture, but it is also the basis of the van Cittert–Zernike theorem in the following section.

3.1.7 van Cittert–Zernike Theorem

The famous van Cittert–Zernike theorem was developed from the work of van Cittert in 1934 and Zernike in 1938. This theorem (Born and Wolf, 1999) formalizes the heuristic discussion in the preceding section, and it is the basis of any attempt to reconstruct an object from interferometer measurements.

Suppose that two apertures are separated by baseline vector $\vec{B}$. Suppose that the source has an intensity distribution on the sky given by $I(\vec{\alpha})$, where $\vec{\alpha}$ is a two-dimensional sky coordinate. Then the *complex degree of coherence* μ is given by

$$\mu(\vec{B}) \quad = \quad \int I(\vec{\alpha}) e^{-ik\vec{B}\cdot\vec{\alpha}} d\vec{\alpha} / \int I(\vec{\alpha}) d\vec{\alpha} \tag{3.43}$$

where $k = 2\pi/\lambda$, and the integrals are over the field of view of the diffraction-limited single-aperture beam, a cone of half-angle $\theta_{\rm tel}$. The *degree of coherence* is the modulus

$$\text{visibility} = |\mu| \tag{3.44}$$

and the phase is the argument

$$\text{phase} = \arg(\mu). \tag{3.45}$$

The inverse relation

$$I(\vec{\alpha}) / \int I(\vec{\alpha}) d\vec{\alpha} \quad = \quad \int \mu(\vec{B}) e^{+ik\vec{B}\cdot\vec{\alpha}} d\vec{B} \tag{3.46}$$

recovers the image from the suite of visibility measurements. In this equation, the integral on the right is over all possible baseline positions.

3.2 Beam Combination in Practice

This section addresses some practical aspects of beam combination, including Michelson's pioneering stellar interferometer, optical configurations for large ground-based interferometers, and multiplexing methods.

3.2.1 Michelson's Stellar Interferometer

An interesting aspect of Michelson's original stellar interferometer is that although it was mounted on a conventional telescope structure, the actual wavefront collecting mirrors were not part of the telescope optics, but rather were a pair of 45° flats riding on an external rail so that the baseline length B could be adjusted to be up to several times larger than the telescope primary itself. The beam combination itself used a smaller baseline B_0 which had used the telescope primary to bring together the separate beams and form superposed images of the star in the focal plane, as shown in Figure 3.4.

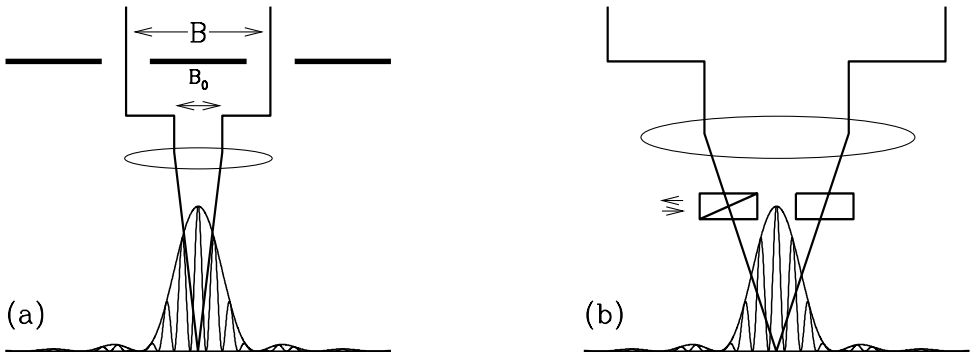

Figure 3.4: (a) Michelson interferometer schematic, showing the external collection baseline B, and the internal combination baseline B_0. The fringe spacing is determined by B_0. (b) Detail schematic of variable-thickness wedge (left) and tilt plate (right) used by Michelson to equalize the optical paths at all wavelengths and to precisely superpose two images of the target star, respectively.

Thus the coherence of the star is measured by the collecting baseline B, so this value governs the visibility of the source. The modulation pattern in the focal plane is set by the combining baseline B_0. This is a sufficiently important distinction that we write it out:

$$B = \text{collection baseline} \tag{3.47}$$

$$B_0 = \text{combination baseline} \tag{3.48}$$

So, with this distinction in mind, and without further derivation, we write the intensity in the focal plane of the interferometer as

$$I_{\rm int}(\theta) = 2I_{\rm tel}(\theta)\,[1 + V_{\rm UD}\cos(2\pi\theta B_0/\lambda)] \tag{3.49}$$

where the B_0–dependent cosine term expresses the modulation of the envelope $I_{\rm tel}$, and where the B–dependent $V_{\rm UD}$ term expresses the degree of modulation

$$V_{\rm UD} = \frac{2J_1(\pi B\theta_{\rm UD}/\lambda)}{\pi B\theta_{\rm UD}/\lambda}. \tag{3.50}$$

So B_0 can be made to be any convenient value. Michelson used $B_0 = 1.14$ m, so the fringe width is $\theta_{\rm int} = \lambda/2B_0 = 0.045$ arcsec. Let us assume that Michelson's viewing eye had an angular resolution $\theta_{\rm eye} = 1.22\lambda/(5\text{ mm}) \simeq 25$ arcsec. To resolve the fringes with his eye he would therefore need additional angular magnification M from the telescope eyepiece, where nominally $M = \theta_{\rm eye}/\theta_{\rm int} \simeq 25/0.045 \simeq 600$ times, and, in fact, that is what he reported.

Two further details are worth noting. First, Michelson inserted a plane-parallel plate of glass in one of the beams, within arm's reach, and tilted it in order to precisely superpose

the two star images, thereby effectively making the wavefronts parallel at the entrance pupil, and compensating for small alignment errors in the relay flats and bending of the structure. (Remember, position in the focal plane corresponds to angle at the incoming wavefront.)

Second, Michelson inserted two opposing glass wedges into the other beam, and slid these past each other so as to give a variable thickness of glass, and thereby compensate for the variable effective thickness of the tilt plate. This compensation ensures that wavefronts from different wavelengths arrive at the same time from both beams, and ensures that the broad band of wavelengths produces sharp fringes across the intensity envelope.

3.2.2 Image-Plane and Pupil-Plane Combination

There are two fundamentally different types of beam combination at the back end of an interferometer, and all ground- and space-based interferometers use one or the other of these methods. Deciding which one to use depends to some extent on the personal style of the designer. In principle, with an ideal instrument (noiseless detector, etc.), the ultimate signal-to-noise ratios from both methods should be identical.

Image-Plane Interferometry

Image-plane interferometry is the method of combining two beams in which each beam is focussed to make an image of the sky, and the images are superposed, so that interference fringes will form across the combined image. This is also called *Fizeau interferometry*, after Fizeau who originally suggested using a two-slit mask across the aperture of a conventional telescope to resolve stellar diameters. It is also the method used by Michelson in his stellar interferometer, as described above.

A generalized image-plane interferometer, for ground-based observations with long baselines, is sketched in Figure 3.5(a). Let ζ be the angle between the baseline vector $\vec{B}$ and the stellar wavefront above the atmosphere. Then a stellar wavefront arrives at one telescope with an external vacuum path difference $z_{\text{ext.}} = B\sin(\zeta)$ compared to its arrival at the other telescope. To compensate, a delay line is introduced into one arm of the interferometer, giving an internal vacuum delay $z_{\text{int.}}$. The phase difference between the two beams is then $\phi = 2\pi\Delta z/\lambda$, where $\Delta z = (z_{\text{ext.}} - z_{\text{int.}})$ is the optical path difference (OPD) between the wavefronts, and the delay line is continuously adjusted to keep this quantity close to zero, as the Earth rotates.

The fringe intensity is displayed as a function of angular position θ in the focal plane, and is given by

$$I_{\text{int}}(\theta) = 2I_{\text{tel}}(\theta)\,[1 + V\cos(2\pi(\theta B_0 + \Delta z)/\lambda)] \tag{3.51}$$

where I_{tel} is the envelope shape, V is the visibility of the star, θB_0 is the fringe modulation

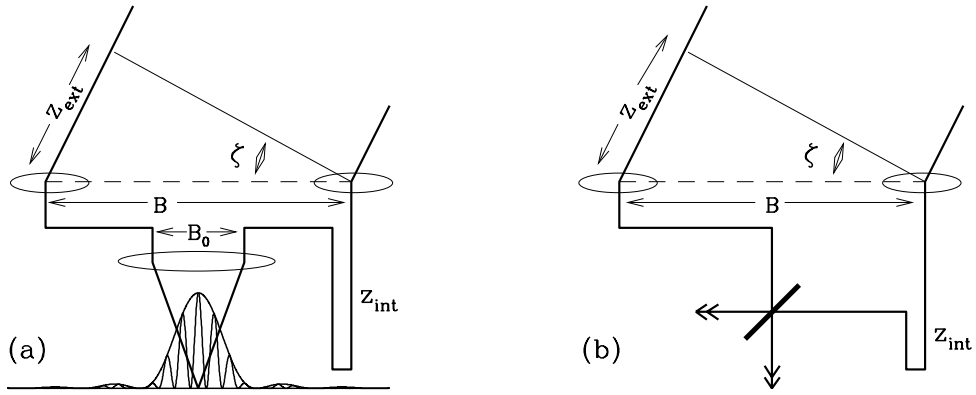

Figure 3.5: (a) Image-plane interferometer schematic,showing the external path difference $z_{\text{ext.}}$, internal path difference $z_{\text{int.}}$, and the non-zero beam-combination baseline B_0. (b) Pupil-plane interferometer schematic, showing a half-silvered beam-combiner plate and zero distance beween combining beams.

term and Δz is the fringe position term. Appropriate magnification and a multi-element detector are used to detect the fringes.

When the delay line is adjusted to give an OPD of $\Delta z = 0$, then the peak intensity of the fringe pattern is centered in the envelope. If the delay line is moved off of the zero OPD position, the envelope will stay fixed but the fringes will move across the envelope. If a finite bandwidth or multiple narrow wavelength bands are present, then at non-zero OPD values the fringe peaks from different wavelengths will be non-coincident, and for large values of OPD the fringes will blur out completely, as described in the section below, under "spectral bandpass."

Spectral dispersion can be used with image-plane interferometry by introducing a prism or grating to disperse the light along the direction of the fringes. Since the fringe spacing is proportional to λ, the dispersed fringes will have a fan-like appearance with the red fringes at the wide end of the fan, and blue at the narrow end.

Pupil-Plane Interferometry

Pupil-plane interferometry is the method of combining two beams in which parallel beams are superposed, using a half-silvered mirror or equivalent, and the two resulting output beams are each focussed on single detector pixels. This is called *Michelson interferometry*, after the eponym's original 1893 interferometer, which showed that the speed of light is independent of the observer's velocity.

A ground-based pupil-plane interferometer is sketched in Figure 3.5(b). In pupil-plane interferometry the combining beams are completely overlapped, so the combining baseline is zero, $B_0 = 0$. The delay-line compensation is identical to that in the preceding case.

If the OPD in Figure 3.5(b) is adjusted to be zero, then by symmetry the overlapped beams emerging from either side of the beam splitter should have equal intensities, since each is the sum of one reflected and one transmitted beam. Another way to express this is to say that the beam splitter has the property that the phase difference between transmitted and reflected beams is exactly $\pi/2$, which we shall prove in the following section entitled "beam splitter phase shift," but which we will simply accept for the moment.

The phase difference between the combined beams is then $\phi = 2\pi\Delta z/\lambda \pm \pi/2$, where $\Delta z \equiv z_{\text{ext.}} - z_{\text{int.}}$ is the OPD as before, and where the additional $\pm\pi/2$ is the beam splitter phase shift.

The fringe intensity follows a similar expression as above, but the $\pi/2$ term changes the $+\cos$ term to $\pm\sin$. Thus as the OPD is varied, the combined beam intensities will vary with opposite signs, giving

$$I_{\text{int}}(t) = 2I_{\text{tel}}\left[1 \pm V \sin(2\pi\Delta z(t)/\lambda)\right]. \tag{3.52}$$

Here we have integrated over angle in the focal plane so that $I_{\text{tel}} = \int I_{\text{tel}}(\theta)d\theta$. We have explicitly assumed that the phase difference is time-modulated, rather than being spatially modulated as in the image-plane case. If the time modulation is a triangle or ramp function, then $\Delta z(t) = vt$ over part of the modulation cycle. The measured amplitude of the time-modulated signal gives the visibility V directly.

Note that the $\pi/2$ term, or equivalently the sine dependence, is very frequently ignored by practitioners and textbooks alike, but it is nevertheless a salient feature of pupil-plane interferometry.

If a finite spectral bandwidth is present, then, just as in the image-plane case, if the OPD is adjusted to be non-zero, the fringe peaks of the different wavelengths will fail to overlap perfectly, and the fringes will blur out, as described in the "spectral bandpass" section below.

Spectral dispersion can be used with pupil-plane interferometry by adding a prism or grating just before detection, so that adjacent wavelengths fall on adjacent detector pixels. The resulting display is called a channel spectrum, because for non-zero path differences the spectrum will be wavelength-modulated by a sinusoidal intensity pattern (opposite in the two output beams) with the appearance of channels in an otherwise smooth spectrum. The depth of modulation at each wavelength gives the visibility directly.

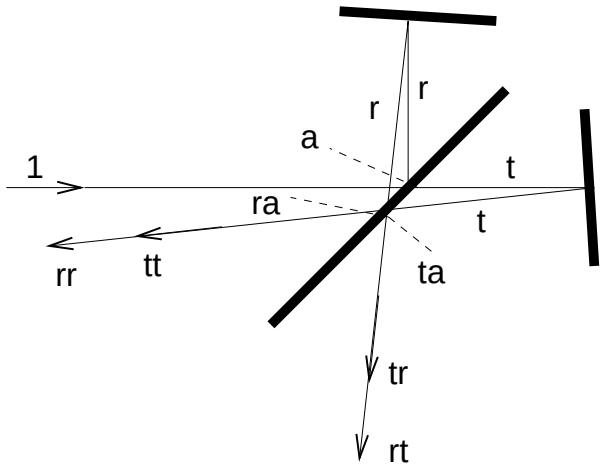

Figure 3.6: Beam splitter experiment showing a unit-intensity beam incident on a thin beam splitter, with two mirrors arranged so as to give equal length arms. The emerging beams are parallel and overlapping. Losses due to absorption or scattering are indicated.

3.2.3 Beam splitter Phase-Shift

In this section we prove the remarkable fact that the phase difference between reflected and transmitted beams from a beam splitter is $\pi/2$.

Suppose that we have a thin, symmetric beam splitter, such as a thin metal layer suspended in space or sandwiched between two identical sheets of glass. Suppose that the relative amplitude of a wavefront reflected from this beam splitter is r with phase shift δ_r, the relative amplitude of the transmitted wavefront is t with phase shift δ_t, and the relative amplitude absorbed or scattered is a.

Suppose that we set up a lab experiment as shown in Figure 3.6. The incident beam has amplitude $A_0 = 1$ from one side of the beam splitter, and zero from the other side.

The incident beam is split into a reflected complex amplitude $re^{i\delta_r}$, a transmitted complex amplitude $te^{i\delta_t}$, and an absorbed amplitude a. The corresponding relative intensities are reflectance $R = |r|^2$, transmittance $T = |t|^2$, and absorptance $A = |a|^2$. Each split beam is then reflected by a perfectly reflecting mirror and returned to the beam splitter with identical delay and phase shift on reflection in each arm; these terms will factor out, so we ignore their effect here to keep the equations uncluttered.

The returned beams then each split again as before, and are partially absorbed as well. The emerging amplitudes are then given by

$$\begin{aligned} A_1 &= re^{i\delta_r}te^{i\delta_t} + te^{i\delta_t}re^{i\delta_r} && (3.53)\\ &= 2rte^{i(\delta_r+\delta_t)} && (3.54)\\ A_2 &= re^{i\delta_r}re^{i\delta_r} + te^{i\delta_t}te^{i\delta_t} && (3.55)\\ &= r^2e^{i2\delta_r} + t^2e^{i2\delta_t} && (3.56) \end{aligned}$$

The absorbed amplitudes are

$$A_3 = a \quad ; \quad A_4 = ra \quad ; \quad A_5 = ta. \tag{3.57}$$

The corresponding intensities of the incident, reflected, and absorbed beams are

$$\begin{aligned} I_0 &= 1 && (3.58)\\ I_1 &= 4RT && (3.59)\\ I_2 &= R^2 + 2RT\cos^2(\delta_r - \delta_t) + T^2 && (3.60)\\ &= R^2 - 2RT + T^2 + 4RT\cos^2(\delta_r - \delta_t) && (3.61)\\ I_3 + I_4 + I_5 &= A + RA + TA && (3.62) \end{aligned}$$

Conservation of energy requires that $I_0 = I_1 + I_2 + I_3 + I_4 + I_5$. Inserting the above values and simplifying, we find that the phase shifts are required to obey

$$\begin{aligned} \cos^2(\delta_r - \delta_t) &= 0 && (3.63)\\ |\delta_r - \delta_t| &= \pi/2 && (3.64) \end{aligned}$$

Thus a thin beam splitter will have a $\pi/2$ phase shift between the reflected and transmitted beams, independent of the reflection, transmission, and absorption in the beam splitter. (I suspect, but have not shown, that this result still applies to finite-thickness non-absorbing beam splitters, but that the result fails for finite-thickness absorbing asymmetric beam splitters.)

By substituting $\delta_r = \delta_t \pm \pi/2$ back into the amplitude equations, it is easy to show that the output beams both have the same phase, $+2\delta_t$, which is interesting, but has no immediate application.

By repeating the entire derivation with unequal arm lengths it is also easy to see that the output beam intensities are complementary, i.e., that the intensities add to a constant value. This result does have great value, because it means that in a real pupil-plane interferometer, the sum of the output intensities can be used to normalize unavoidable intensity fluctuations due to atmospheric or other perturbations. Since these fluctuations often exceed photon-counting (Poisson) fluctuations on relatively bright stars, the technique of intensity normalization is a valuable tool for maximizing the observed signal-to-noise ratio.

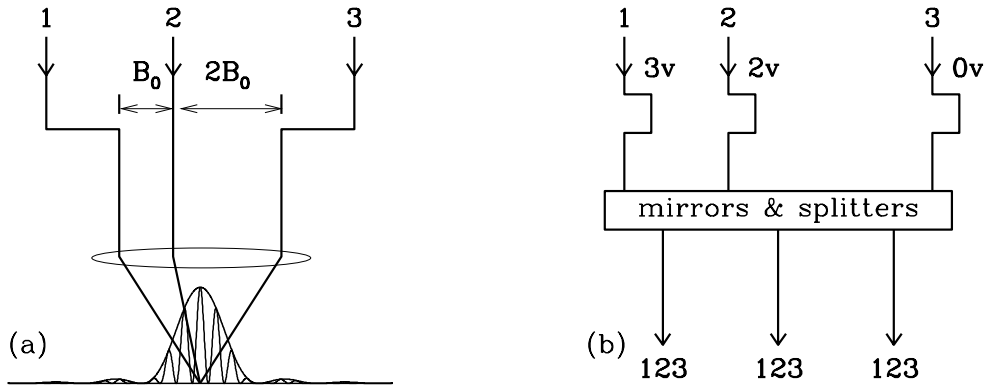

Figure 3.7: (a) A multiplexing scheme for a pupil-plane interferometer showing three combination baselines $1B_0, 2B_0, 3B_0$ for three beams. The input baselines B_{ij} can be arbitrary, and are independent of the output baselines. (b) A multiplexing scheme for an image-plane interferometer, showing three combination Doppler-shift velocities $3v, 2v, 0v$ for three beams. Here too, the input baselines can be arbitrary.

3.2.4 Multiplexing Three or More Apertures

For N apertures or telescopes, there are $N(N-1)/2$ baselines among the apertures. The N beams can be combined in $N(N-1)/2$ pairs, with each pair being detected as discussed above. However the N beams can also be combined all at once on a single detector pixel, if care is taken to encode each different pair of telescope beams with a different modulation frequency so that the component pairs can be extracted in post-processing.

There are two advantages to multiplexing. First, in the case where detector noise is greater than photon shot noise, it is advantageous to put the largest possible signal on the fewest possible detector elements. Second, in the case where phase-closure measurements are being made (which requires three or more telescopes), it is advantageous to have all the beams traversing the same paths as much as possible, to avoid unmeasured path changes in the optics due to temperature changes, etc., and this leads naturally to having all beams superposed.

A multiplexing image-plane interferometer can be made by arranging the output beam separations in a minimum redundancy array, so that the spatial frequencies in the image plane all have different values. This is illustrated in Figure 3.7(a) where the separations are B_0 and $2B_0$, so that the squared visibilities at each of the output spatial frequencies are in the proportions $f_{12} \sim 1$, $f_{23} \sim 2$, and $f_{31} \sim 3$, for example.

The power spectral density of the spatial intensity distribution is defined as

$$PSD = |FFT(\text{fringe pattern})|^2. \tag{3.65}$$

A plot of the PSD against spatial frequency will give distinct peaks at each of these frequencies, proportional to the image content at the corresponding external baselines B_{12}, B_{23},and B_{31}. Alternatively a two-dimensional $\vec{B_0}$ pattern could be used, and a two-dimensional FFT extraction performed.

A multiplexing pupil-plane interferometer can be made by arranging the input beam delays to have different delay-line speeds, so that the Doppler shifts of the combined beams are each different. Figure 3.7(b) illustrates this with delay line velocities in the ratios $v_1 \sim 3$, $v_2 \sim 2$, $v_3 \sim 0$, so that the temporal PSD of any of the combined beams will contain distinct peaks at the output temporal frequencies in the ratios $f_{12} \sim 1$, $f_{23} \sim 2$, and $f_{31} \sim 3$, and the power in each peak is proportional to the visibility squared.

3.3 Visibility Loss Effects

High-quality measurements require that the observer minimize and calibrate the instrumental losses of visibility. Some of these effects can be minimized by proper design of the interferometer, and some by operation; all effects can be calibrated out of the data, in principle. A recent examination of instrumental sources of visibility loss for the IOTA interferometer, including more effects than listed here, is given by Porro *et al.* (1999).

The Strehl ratio S is defined as the ratio of (a) the measured peak intensity $I_{\rm meas}(\max)$ of an image formed by a real optical system, including optical aberrations, and frequently including atmospheric seeing; and (b) the idealized peak intensity $I_{\rm ideal}(\max)$ of an image formed by an ideal optical system, including only the effects of diffraction, and not including atmospheric seeing. Thus the Strehl ratio for images is

$$S = I_{\rm meas}(\max)/I_{\rm ideal}(\max). \tag{3.66}$$

The Strehl ratio concept is also applicable to the fringe modulation in an interferogram, for either spatially or temporally displayed fringes. By analogy we write the Strehl ratio for fringes as

$$S = V_{\rm meas}(\max)/V_{\rm ideal}(\max). \tag{3.67}$$

To estimate the combined effect of different sources of visibility, or of Strehl ratio, from the star and the instrument, the general practice is simply to multiply the various factors together, because we assume that they are all independent. Although this cannot be strictly valid, it is a very good approximation for small perturbations.

Atmospheric fluctuations can also cause visibility losses, and these can be more troublesome than instrumental losses because potentially they are larger in magnitude and variable in time; these effects are discussed by Quirrenbach in Chapter 5.

The results in this section will be stated without derivation; however, using the principles outlined above, the derivations could be supplied by the reader.

3.3.1 Spectral Bandpass

All stellar measurements use a finite range of wavelengths, or bandwidth. Any fringe packet, whether it is displayed spatially or temporally, will suffer a reduction in modulation amplitude at the edges of the packet, where the different wavelengths will produce opposing peaks and valleys. Suppose that the spectral bandpass is rectangular, and has a center and full-width at half-maximum (FWHM) of $(\lambda, \Delta\lambda)$ wavelengths, or $(\sigma, \Delta\sigma)$ wavenumbers, where $\sigma = 1/\lambda$ and $\Delta\sigma = \Delta\lambda/\lambda^2$ and where $\Delta\lambda/\lambda = \Delta\sigma/\sigma$.

In this case the visibility decreases with distance from the zero path-difference point according to

$$V_{\rm bandpass}(\Delta z) = \frac{\sin(\pi \Delta z \Delta\sigma)}{\pi \Delta z \Delta\sigma} \tag{3.68}$$

$$= \frac{\sin(\pi \Delta z \Delta\lambda/\lambda^2)}{\pi \Delta z \Delta\lambda/\lambda^2}. \tag{3.69}$$

Here the path difference is given by $\Delta z = z_{\rm ext.} - z_{\rm int.}$ as discussed in the previous section.

The first zero of this function is at $\Delta z \Delta\sigma = 1$, from which we find that the number of fringes $N_{\rm bandpass}$ between envelope zero-crossings in a finite-bandpass wave packet is

$$N_{\rm bandpass} = 2\lambda/\Delta\lambda. \tag{3.70}$$

In the general case there is a Fourier-transform relation between the bandpass shape and the fringe packet shape. This is illustrated in Figure 3.8, for the case of a real K-band filter, which is approximately rectangular. Note the sidelobe ringing, which results from aliased beating of the various wavelength fringe patterns outside the main lobe.

In Figure 3.8(a) we measure the filter width to be $\Delta\lambda \simeq 0.40$ μm, so in the fringe packet we expect $N_{\rm bandpass} \simeq 11$ fringes, and this is in fact about what we see in corresponding fringe packet in Figure 3.8(b).

In a pupil-plane interferometer, as the delay line is scanned through the white-light point, the pattern in Figure 3.8(b) is exactly the observed modulation of intensity vs time. In the extreme case of no spectral filtering, the wave packet will tend towards a single spike delta-function. At the other extreme of a very narrow filter, the wave packet will be very many wavelengths wide.

3.3.2 Wavefront Tilt

If two wavefronts of width D are tilted by an angle α, then the interference pattern will be smeared and visibility reduced. The one-dimensional visibility factor from this effect is

$$V_{\rm tilt} = \frac{\sin(\pi D\alpha/\lambda)}{\pi D\alpha/\lambda} \tag{3.71}$$

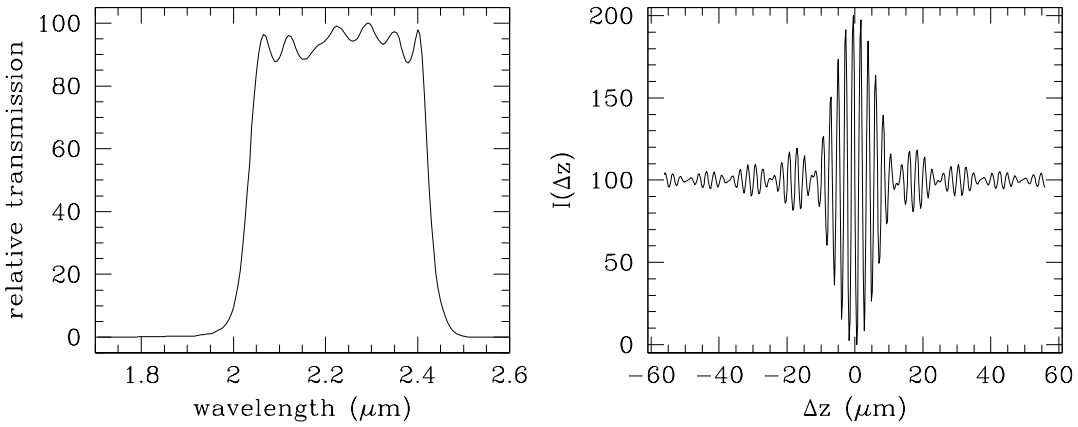

Figure 3.8: (a) Measured K filter transmission profile. (b) Calculated wave packet shape of a pupil-plane interferometer temporal scan through the zero-path-difference point, with the K filter shown. Note that the wave packet is sine-modulated, not cosine-modulated, as is appropriate for an ideal beam splitter.)

for a slit or rectangular aperture. The visibility factor from a two-dimensional circular aperture of diameter D is

$$V_{\rm tilt} = \frac{2J_1(\pi D\alpha/\lambda)}{\pi D\alpha/\lambda}. \tag{3.72}$$

If you wish to have $V_{\rm tilt} > 0.90$, say, then you need to be sure that the wavefronts combine at an angle $\alpha < 0.3\lambda/D$. For example, in a pupil- or image-plane interferometer, where in either case a star image will be formed, this amounts to a star-image-overlap criterion of about 25% of a diffraction-limited spot.

Likewise, a star tracker system, which directly controls wavefront tilt on a continuous basis, will have to peform at least as well, i.e., to 25% of the telescope's diffraction limit, to ensure that the measured visibilities do not fluctuate appreciably. The star-tracker system is thus a crucial part of an interferometer, and it can be a challenging task to achieve an optimum design.

3.3.3 Intensity Mismatch

If the relay optics fail to perfectly overlap the beams from each telescope, or if the beam combiner has unequal reflection and transmission factors, or if the combined beams come from different diameter telescopes and therefore have different intensities in the overlap region, then we will have a reduction in visibility from any of these factors. Let the intensity ratio between one beam and another be written as $I_1/I_2 = \rho$. The visibility factor from this effect is

$$V_{\rm mismatch} = \frac{2}{\rho^{+1/2} + \rho^{-1/2}}. \tag{3.73}$$

This is a relatively tolerant effect. The reason is that amplitudes vary as the square root

of intensity, so two amplitudes will always be relatively closer in numerical value than the corresponding two intensities.

For example, if the beam combiner has an intensity reflection $R = 60\%$ and a transmission $T = 40\%$, then $\rho = 1.5$, so $V_{\rm mismatch} \simeq 0.98$, and there is very little loss of visibility.

3.3.4 Optical Surface Figure Errors

If the combining wavefronts each have a root-mean-square (rms) perturbation of δ with respect to a perfect wavefront, and if the perturbations are randomly distributed across the wavefront, and uncorrelated between the two wavefronts, then the Strehl ratio and the fringe visibility will be degraded (Born and Wolf, 1999) according to

$$V_{\rm surfaces} \simeq e^{-(2\pi\delta/\lambda)^2}. \tag{3.74}$$

If there are N surfaces with rms of δ_0 each, then

$$\delta \simeq N^{1/2}\delta_0. \tag{3.75}$$

For example, if we use a common optical polishing criterion which balances quality and cost, each flat mirror will have a peak-to-valley (pv) surface flatness of $\lambda_0/20$, where $\lambda_0 = 0.632$ μm is the laser measurement wavelength in the optical shop. Experience with several measured mirrors suggests that pv and rms are related by a factor pv/rms $\simeq 5.5$. Suppose these mirrors are used at an average angle of incidence of $45^{\rm o}$. The reflected wavefront will be two times worse than the mirror itself. Combining these factors, we get the wavefront rms at $45^{\rm o}$ incidence from a single $\lambda_0/20$ pv mirror as

$$\delta_0 = \frac{2(\lambda_0/20)}{5.5\cos(45^{\rm o})} \tag{3.76}$$

$$= \lambda_0/39 \tag{3.77}$$

for each reflection. Suppose that there are $N = 14$ mirrors in a typical ground-based interferometer. The net wavefront rms after N reflections will then be $\delta \simeq \lambda_0/10$, and the visibility from this alone will be $V_{\rm surfaces} \simeq e^{-(\pi/5)^2(\lambda_0/\lambda)^2}$. If the test and operating wavelengths are the same, then we find $V_{\rm surfaces} \simeq 0.67$. This is a significant loss, and it shows that the cumulative effect of even rather good optical surfaces can strongly affect an interferometer. If the operating wavelength is longer, then the visibility is improved. See, e.g., Porro *et al.* (1999) for a complete discussion.

3.3.5 Polarization Effects

We tend to ignore polarization, perhaps because our eyes are not sensitive to it, but anyone with the type of polarized sunglasses that you could buy at one time will tell you that reflected light, from the sky or pavement or automobile hood, can be highly polarized. It should be no surprise then that polarization can reduce fringe visibility, as we now show.

Suppose we consider a typical flat mirror which reflects light incident at 45° from the normal. The electric vector components which are perpendicular and parallel to the plane of incidence are called the s and p components respectively. A typical overcoated silver mirror will cause the difference between the s and p phases to change by about 30° in the visible, with the change at other wavelengths varying roughly as λ^{-1}, i.e., less in the infrared.

If an interferometer can be built so that the reflections in each arm follow the same sequence of changes of direction, and if the corresponding mirrors at each reflection are of the same type, then both beams will experience the same phase shifts, and the respective s and p components will combine independently in the focal plane and produce identical fringe packets. This is the principle behind the layout of the IOTA interferometer (Traub, 1988), for example.

However if the sequence of reflections is different, or the mirrors are not the same, then $s - p$ differences can occur. Suppose that the $s - p$ shift between the two beams is $\phi_{\rm sp}$. Then the interferogram will have a visibility term $V_{\rm pol}$, where

$$V_{\rm pol} = |\cos(\phi_{\rm sp}/2)|. \tag{3.78}$$

Note that if $\phi_{\rm sp} = \pi$, then the interferograms from each polarization will be modulated such that the peaks of one occur in the valleys of the other, and the net modulation will be zero. Thus it is possible to make the net interference effect disappear completely! See Traub (1988) for more details.

3.4 Visibility Enhancement Methods

Among the many fascinating methods that have been invented to enhance the instrumental visibility, we briefly discuss four which are of particular interest for current ground- and space-based interferometers: adaptive optics, single-mode fiber optics, single-mode integrated optics, and nulling.

3.4.1 Adaptive Optics

If an adaptive optics (AO) system is used at a telescope the distortion of an input wavefront can be measured and corrected in real time. A compensating distortion is applied to a mirror, so that the resulting wavefront is (in principle) perfectly flat. The measurement is done using either a natural guide star (NGS) or laser guide star (LGS) as a wavefront reference. The technology has been dramatically demonstrated at large ground-based telescopes, where a typical image-width reduction of roughly a factor of 10, and a central intensity increase of a factor of $10^{1.5}$ can be achieved. To date, the technique has not yet been applied to an interferometer, but it will be required when interferometer mirror diameters much exceeding 1 m are used.

The advantage of AO is that large telescope diameters can be used, independent of the atmospheric coherence length. Two references are Roddier (1999) and Hardy (1998).

3.4.2 Fiber Optics

Single-mode fiber optics may be used within an interferometer (a) to select essentially the plane-wave part of a wavefront, (b) to split a guided wave into any desired intensity ratio, and (c) to interferometrically combine two guided waves. Wavelengths longer than the cutoff wavelength will excite a single electromagnetic mode of a fiber waveguide; the cutoff wavelength is a simple function of core radius, core index, and cladding index. Fiber couplings are formed by arranging the cores of two fibers to run parallel to each other with roughly one core diameter thickness of cladding material between the cores. Under these conditions the core excitation hops periodically between one core and the other. By adjusting the interaction length one can achieve any desired degree of transfer, including roughly 50:50, at a particular wavelength.

The advantage of single-mode fibers is that when only the plane-wave part of the wavefront is used, the fluctuations in visibility due to random atmospheric warping of the wavefront are dramatically reduced. Typical visibility uncertainties go from 5% with a classical beam combiner to 0.5% with a fiber-optic combiner. Atmospheric effects blur the image and therefore reduce the intensity coupled into each fiber at any given instant. Thus the flux in each fiber must be monitored. This can be easily done by tapping off a portion of the flux with a coupler. Two references are Coudé du Foresto *et al.* (1997) and Delage and Reynaud (2000).

3.4.3 Integrated Optics

The methods of integrated optics (IO) allow single-mode waveguides to be manufactured in-situ on the surface of a plane glass substrate, using integrated-circuit techniques, with all of the advantages of fiber optics, plus the advantage of small size and reduced cost of production. IO will be tested at ground-based interferometers in the near future. Recent progress in this field is described by Malbet *et al.* (1999) and Haguenauer *et al.* (2000).

3.4.4 Nulling

Nulling interferometry is a technique in which a phase shift of π is added to one wavefront segment, so that when it interferes with another segment of the same wavefront, perfect cancellation is achieved on-axis. Thus a bright central star can be dimmed by many orders of magnitude relative to the surrounding off-axis material, such as a planetary system.

There are several techniques which can be used to create the π phase shift. Note that this is not the same as moving a mirror 1/4 wavelength, because for other wavelengths the phase

shift is different. One technique is to use roof reflectors (pairs of mirrors at 90°) to achieve a reversal of sign of the electric vector. Another technique is to introduce a precise thickness of glass whose index will act to retard all wavelengths by very nearly one-half wavelength. Mirror and lens combinations can also be used. To date broad-band nulling of 1 part in 10^4 in the visible has been achieved, using pairs of roof reflectors. Nulling interferometry is discussed in Serabyn in Chapter 16. Four further references are Bracewell and MacPhie (1979), Hinz *et al.* (1998), and Serabyn *et al.* (1999).

References

M. Born and E. Wolf, *Principles of Optics*, 7 (expanded) edn. (Cambridge, UK: Cambridge University Press, 1999).

R.N. Bracewell and R.H. MacPhie, "Searching for nonsolar planets," Icarus **38**, 136–147 (1979).

L. Delage and F. Reynaud, "Analysis and control of polarization effects on phase closure and image acquisition in a fibre-linked three-telescope stellar interferometer," J. Opt. A: Pure Appl. Opt. **2**, 147–153 (2000).

V. Coudé du Foresto, S. Ridgway, and J.-M. Mariotti, "Deriving object visibilities from interferograms obtained with a fiber stellar interferometer," Astron. Astrophys. Sup. Ser. **121**, 379–392 (1997).

P. Haguenauer, J.P. Berger, K. Rousselet-Perraut, P. Kern, F. Malbet, I. Schanen-Duport, and P. Benech, "Integrated optics for astronomical interferometry. III. Optical validation of a planar optics two-telescope beam combiner," Appl. Opt. **39**, 2130–2139 (2000).

J.W. Hardy, *Adaptive Optics for Astronomical Telescopes* (Oxford, UK: Oxford University Press, 1998).

P.M. Hinz, J.R.P. Angel, W.F. Hoffmann, D.W. McCarthy Jr, P.C. McGuire, Matt Cheselka, J.L. Hora, and N.J. Woolf, "Imaging circumstellar environments with a nulling interferometer," Nature **395**, 251–253 (1998).

F. Malbet, P. Kern, I. Schanen-Duport, J.-P. Berger, K. Rousselet-Perraut, and P. Benech, "Integrated optics for astronomical interferometry. I. Concept and astronomical applications," Astron. Astrophys. Sup. Ser. **138**, 135–145 (1999).

I.L. Porro, W.A. Traub, and N.P. Carleton, "Effect of telescope alignment on a stellar interferometer," Appl. Opt. **38**, 6055–6067 (1999).

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2 edn. (Cambridge, UK: Cambridge University Press, 1992).

F. Roddier, ed., *Adaptive Optics in Astronomy* (Cambridge, UK: Cambridge University Press, 1999).

D.J. Schroeder, *Astronomical Optics*, 2 edn. (San Diego, CA: Academic Press, 2000).

E. Serabyn, J.K. Wallace, G.J. Hardy, E.G.H. Schmidtlin, and H. Nguyen, "Deep nulling of visible laser light," Appl. Opt. **38**, 7128–7132 (1999).

W.A. Traub, "Polarization effects in stellar interferometers," in *High Resolution Imaging by Interferometry,* F. Merkle, ed. Proc. ESO Conf. **29**, 1029–1038 (1988).